Kourtesis P, Ledley RS, Suzek BE, Vinayaka CR, Zhang J, Barker WC, The Protein Information Resource, Nucleic Acids Res. 2003 Jan 1;31(1):345-7.); GenBank (ncbi.nlm.nih.gov); PDB (H. M. Berman, T. Battistuz, T. N. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, Z. Feng, G. L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J. D. Westbrook and C. Zardecki, The Protein Data Bank, Acta Cryst. (2002). D58, 899-907) and BIND (Bader, et al., Nucleic Acids Res 29(1):242-245. (2001)).

Please replace the paragraph on page 14, lines 14-21 with the following paragraph.

Similarly, structural alignment of structurally related proteins can be done to generate sequence alignments. There are a wide variety of such structural alignment programs known. See for example VAST from the NCBI (Gibrat, et al., Curr Opin Struct Biol 6(3):377-385. (1996)); SSAP (Orengo and Taylor, Methods Enzymol 266(617-635 (1996)) SARF2 (Alexandrov, Protein Eng 9(9):727-732. (1996)) CE (Shindyalov and Bourne, Protein Eng 11(9):739-747. (1998)); (Orengo et al., Structure 5(8):1093-108 (1997); Dali (Holm et al., Nucleic Acid Res. 26(1):316-9 (1998), all of which are incorporated by reference). These structurally-generated sequence alignments can then be examined to determine the observed sequence variations.

Please replace the paragraph on page 14, lines 22-32 with the following paragraph.

Primary libraries can be generated by predicting secondary structure from sequence, and then selecting sequences that are compatible with the predicted secondary structure. There are a number of secondary structure prediction methods, including, but not limited to, threading (Bryant and Altschul, Curr Opin Struct Biol 5(2):236-244. (1995)), Profile 3D (Bowie, et al., Methods Enzymol 266(598-616 (1996); MONSSTER (Skolnick, et al., J Mol Biol 265(2):217-241. (1997); Rosetta (Simons, et al., Proteins 37(S3):171-176 (1999); PSI-BLAST (Altschul and Koonin, Trends Biochem Sci 23(11):444-447. (1998)); Impala (Schaffer, et al., Bioinformatics 15(12):1000-1011. (1999)); HMMER (McClure, et al., Proc Int Conf Intell Syst Mol Biol 4(155-164 (1996)); Clustal W Higgins D., Thompson J., Gibson T.Thompson J.D., Higgins D.G., Gibson T.J.(1994).CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting,position-specific gap penalties and weight matrix choice.Nucleic Acids Res. 22:4673-4680; BLAST (Altschul, et al., J Mol Biol 215(3):403-410. (1990)), helix-coil transition theory (Munoz and Serrano, Biopolymers 41:495, 1997), neural networks, local structure alignment and others (e.g., see in Selbig et al., Bioinformatics 15:1039, 1999).

Specification Amendments to Update Status of Patent Reference:

Please replace the paragraph on page 2, lines 10-12 with the following paragraph.

In particular, U.S.S.N.s 60/061,097, 60/043,464, 60/054,678, 09/127,926, now US Patent No. 6,269,312and PCT US98/07254 describe a method termed "Protein Design Automation", or PDA, that utilizes a number of scoring functions to evaluate sequence stability.

Please replace the paragraph on page16, lines 17-31 with the following paragraph.

In a preferred embodiment, the computational method used to generate the primary library is Protein Design Automation (PDA), as is described in U.S.S.N.s 60/061,097, 60/043,464, 60/054,678, 09/127,926, now US Patent No. 6,269,312, and PCT US98/07254, all of which are expressly incorporated herein by reference. Briefly, PDA can be described as follows. A known protein structure is used as the starting point. The residues to be optimized are then identified, which may be the entire sequence or subset(s) thereof. The side chains of any positions to be varied are then removed. The resulting structure consisting of the protein backbone and the remaining sidechains is called the template. Each variable residue position is then preferably classified as a core residue, a surface residue, or a boundary residue; each classification defines a subset of possible amino acid residues for the position (for example, core residues generally will be selected from the set of hydrophobic residues, surface residues generally will be selected from the hydrophilic residues, and boundary residues may be either). Each amino acid can be represented by a discrete set of all allowed conformers of each side chain, called rotamers. Thus, to arrive at an optimal sequence for a backbone, all possible sequences of rotamers must be screened, where each backbone position can be occupied either by each amino acid in all its possible rotameric states, or a subset of amino acids, and thus a subset of rotamers.

Please replace the paragraph on page 18, lines 1-4 with the following paragraph.

As outlined in U.S.S.N. 09/127,926, now US Patent No. 6,269,312, the protein backbone (comprising (for a naturally occurring protein) the nitrogen, the carbonyl carbon, the α-carbon, and the carbonyl oxygen, along with the direction of the vector from the α-carbon to the β-carbon) may be altered prior to the computational analysis, by varying a set of paramcters called supersecondary structure parameters.

Please replace the paragraph beginning on page 19, line 29, ending on page 20 at line 6 with the following paragraph.

The classification of residue positions as core, surface or boundary may be done in several ways, as will be appreciated by those in the art. In a preferred embodiment, the classification is done via a visual scan of the original protein backbone structure, including the side chains, and assigning a classification based on a subjective evaluation of one skilled in the art of protein modeling. Alternatively, a preferred embodiment utilizes an assessment of the orientation of the Cα-Cβ vectors relative to a solvent accessible surface computed using only the template Cα atoms, as outlined in U.S.S.N.s 60/061,097, 60/043,464, 60/054,678, 09/127,926, now US Patent No. 6,269,312, and PCT US98/07254. Alternatively, a surface area calculation can be done.

Please replace the paragraph on page 21, lines 9-21 with the following paragraph.

Once the group of potential rotamers is assigned for each variable residue position, processing proceeds as outlined in U.S.S.N. 09/127,926, now US Patent No. 6,269,312, and PCT US98/07254. This processing step entails analyzing interactions of the rotamers with each other and with the protein backbone to generate optimized protein sequences. Simplistically, the processing initially comprises the use of a number of scoring functions to calculate energies of interactions of the rotamers, either to the backbone itself or other rotamers. Preferred PDA scoring functions include, but are not limited to, a Van der Waals potential scoring function, a hydrogen bond potential scoring function, an atomic solvation scoring function, a secondary structure propensity scoring function and an electrostatic scoring function. As is further described below, at least one scoring function is used to score each position, although the scoring functions may differ depending on the position classification or other considerations, like favorable interaction with an α-helix dipole. As outlined below, the total energy which is used in the calculations is the sum of the energy of each scoring function used at a particular position, as is generally shown in Equation 1:

Please replace the paragraph beginning on page 21 at line 28, ending on page 22, line 12 with the following paragraph.

As outlined in U.S.S.N.s 60/061,097, 60/043,464, 60/054,678, 09/127,926, now US Patent No. 6,269,312, and PCT US98/07254, any combination of these scoring functions, either alone or in combination, may be used. Once the scoring functions to be used are identified for each variable position, the preferred first step in the computational analysis comprises the determination of the interaction of each possible rotamer with all or part of the remainder of the protein. That is, the energy of interaction, as measured by one or more of the scoring functions, of each possible rotamer at each variable residue position with either the backbone or other rotamers, is calculated. In a preferred embodiment, the interaction of each rotamer with the entire remainder of the protein, i.e. both the entire template and all other rotamers, is done. However, as outlined above, it is possible to only model a portion of a protein, for example a domain of a larger protein, and thus in some cases, not all of the protein need be considered. The term "portion", as used herein, with regard to a protein refers to a fragment of that protein. This fragment may range in size from 10 amino acid residues to the entire amino acid sequence minus one amino acid. Accordingly, the term "portion", as used herein, with regard to a nucleic refers to a fragment of that nucleic acid. This fragment may range in size from 10 nucleotides to the entire nucleic acid sequence minus one nucleotide.

Please replace the paragraph on page 24, lines 4-6 with the following paragraph.

Once the singles and doubles energies are calculated and stored, the next step of the computational processing may occur. As outlined in U.S.S.N. 09/127,926, now US Patent No. 6,269,312, and PCT US98/07254, preferred embodiments utilize a Dead End Elimination (DEE) step, and preferably a Monte Carlo step.

Please replace the paragraph on page 27, lines 11-22 with the following paragraph.

In a preferred embodiment when scoring is used, although this is not required, the primary library comprises the globally optimal sequence in its optimal conformation, i.e. the optimum rotamer at each variable position. That is, computational processing is run until the simulation program converges on a single sequence which is the global optimum. In a preferred embodiment, the primary library comprises at least two optimized protein sequences. Thus for example, the computational processing step may eliminate a number of disfavored combinations but be stopped prior to convergence, providing a library of sequences of which the global optimum is one. In addition, further computational analysis, for example using a different method, may be run on the library, to further eliminate sequences or rank them differently. Alternatively, as is more fully described in U.S.S.N.s 60/061,097, 60/043,464, 60/054,678, 09/127,926, now US Patent No. 6,269,312, and PCT US98/07254, the global optimum may be reached,